Семинар 2

Задача 1

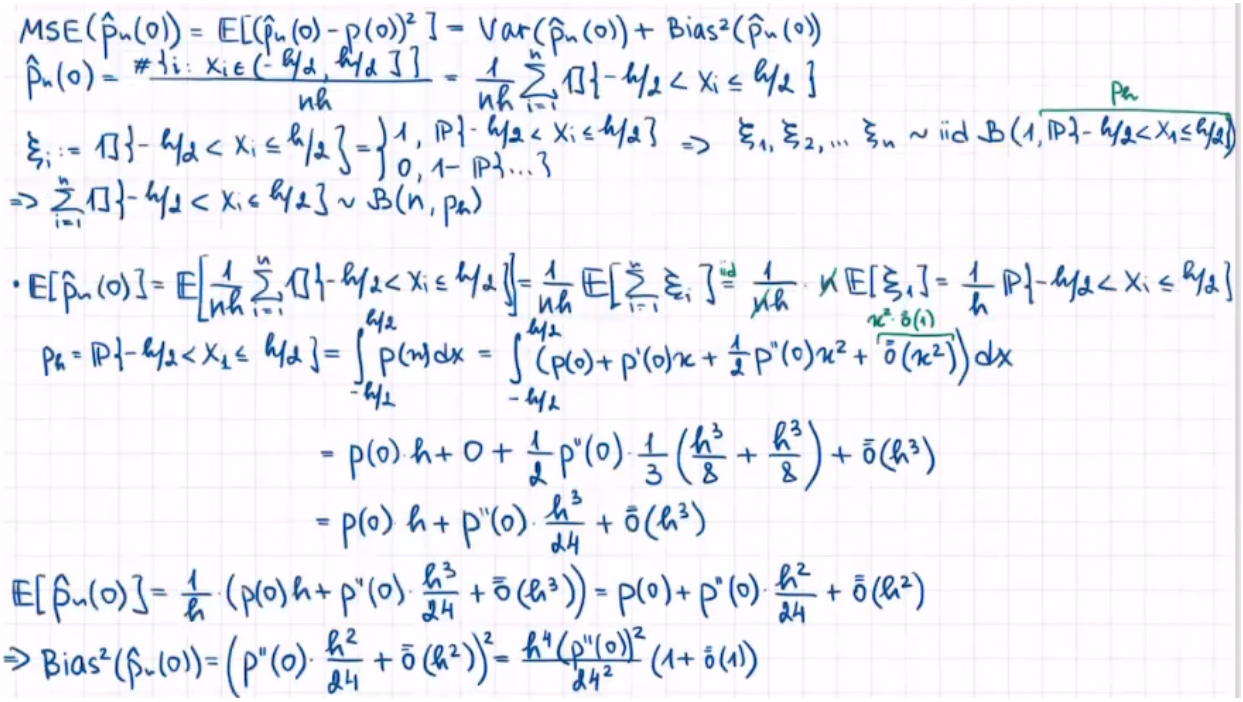
T1 Дана выборка $x_1, \ldots x_n$ из неизвестного абсолютного непрерывного распределения. Для оценивания плотности в точке 0 используется оценка

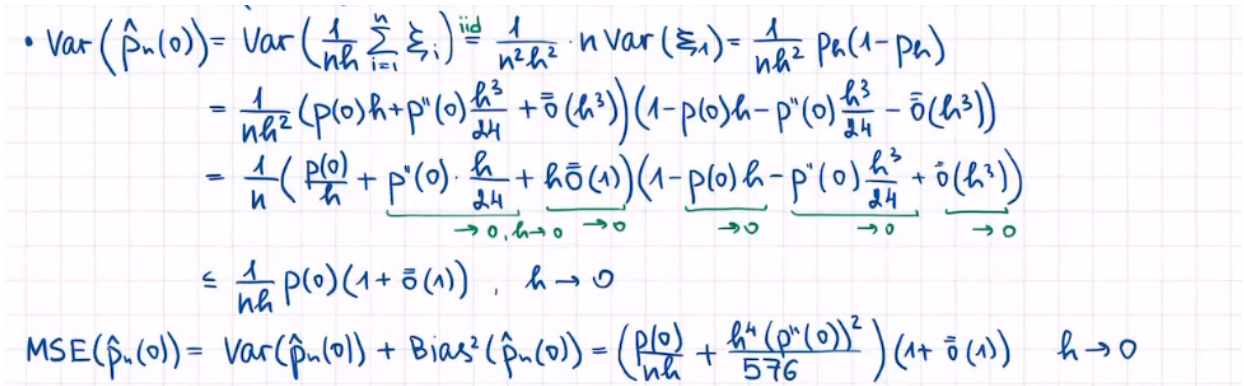
$$\hat{p}_n(0) = \frac{\sharp\{i : x_i \in (-h/2, h/2]\}}{nh}.$$

(i) Докажите, что

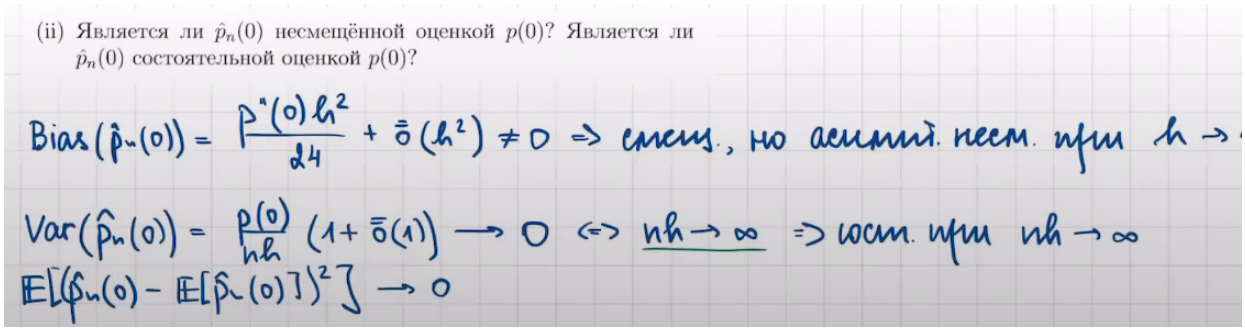
$$MSE(\hat{p}_n(0)) := \mathbb{E}\left[\left(\hat{p}_n(0) - p(0)\right)^2\right] = \left(\frac{(p''(0))^2}{576}h^4 + \frac{p(0)}{nh}\right)(1 + o(1)), \quad n \to \infty,\ h \to 0.$$

Запишем формулу bias-variance tradeoff:





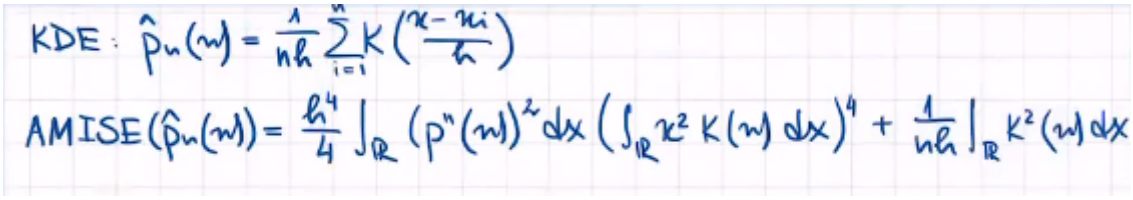
(ii) Является ли $\hat{p}_n(0)$ несмещённой оценкой $p(0)$? Является ли $\hat{p}_n(0)$ состоятельной оценкой $p(0)$?

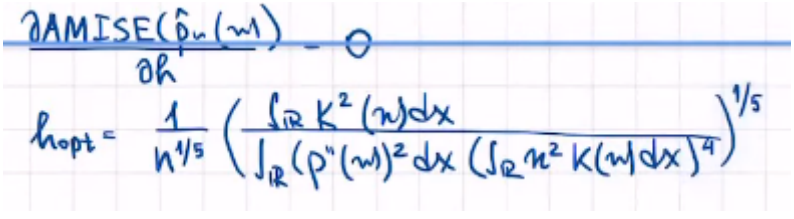


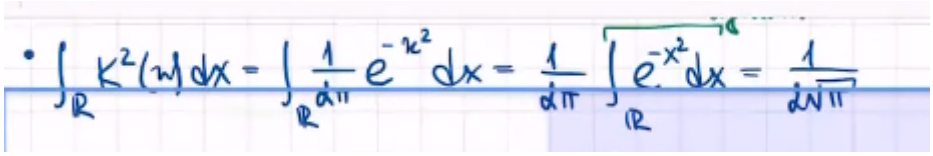
Задача 2

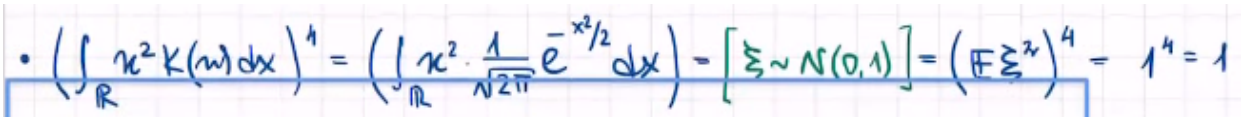
T2 Допустим, что дана выборка из нормального распределения с нулевым средним и дисперсией σ^2. Вычислите значение параметра bandwidth, минимизирующее AMISE (asymptotic mean integrated squared error) ядерной оценки плотности, построенной на основе гауссовского ядра

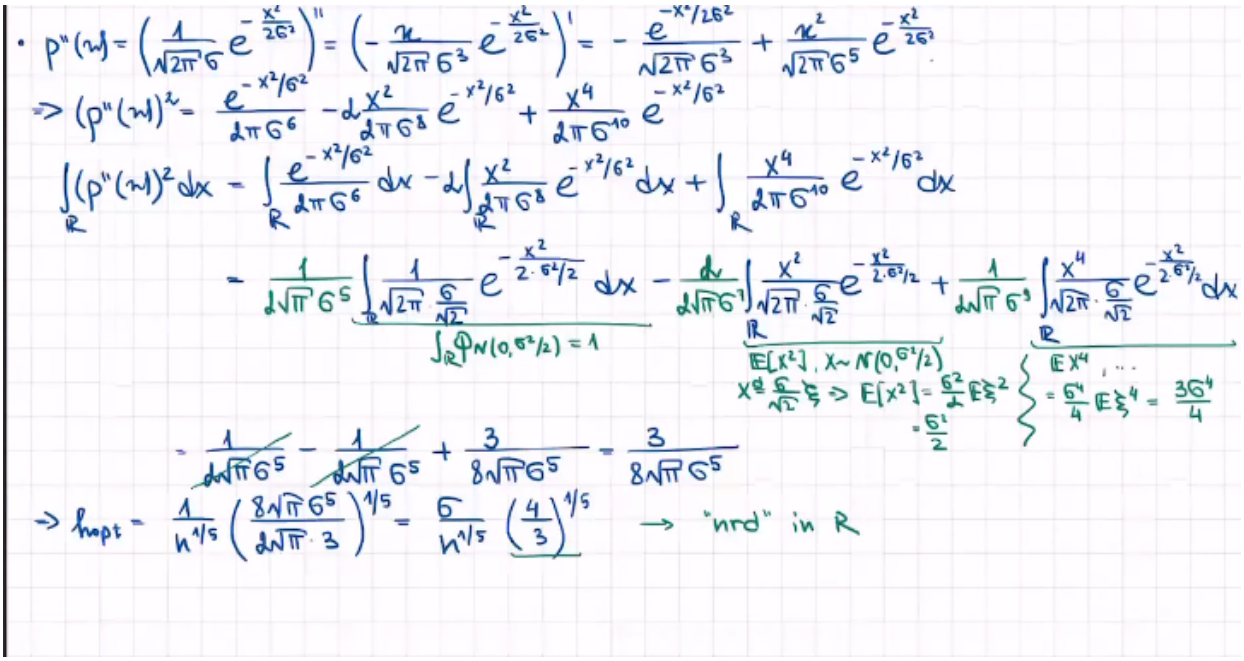
$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$



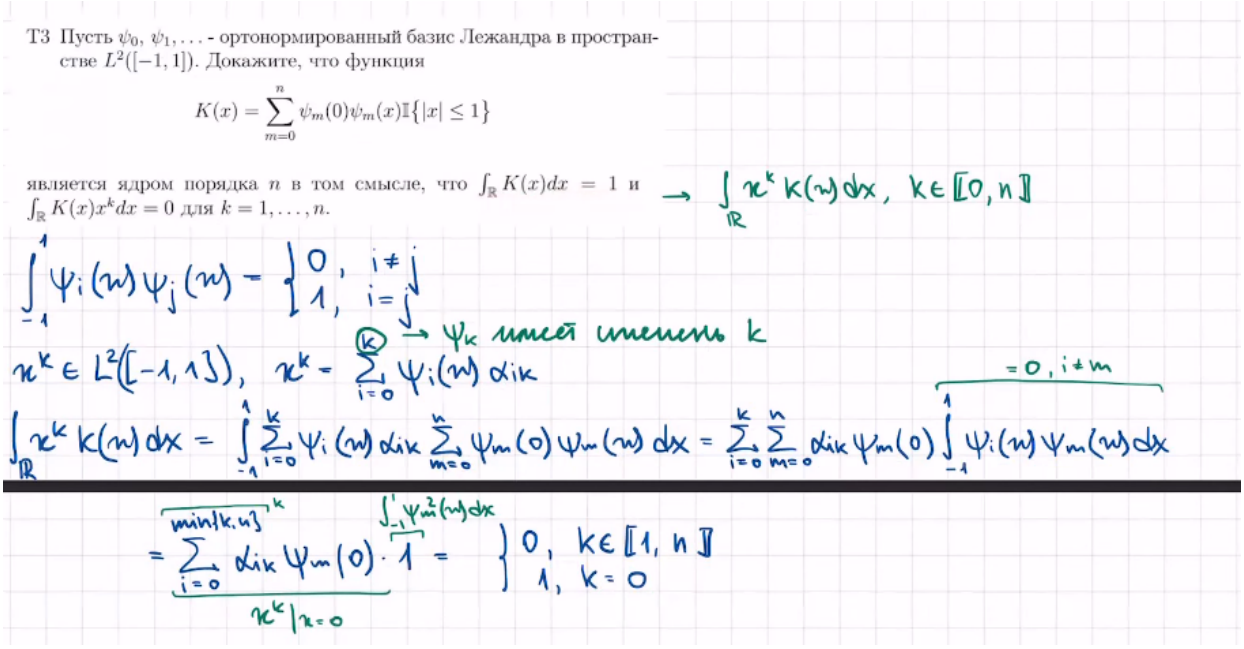








Задача 3



Т3 Пусть $\psi_0, \psi_1, \ldots$ - ортонормированный базис Лежандра в пространстве $L^2([-1,1])$. Докажите, что функция

$$K(x) = \sum_{m=0}^{n} \psi_m(0)\psi_m(x)\mathbb{I}\{|x| \le 1\}$$

является ядром порядка n в том смысле, что $\int_{\mathbb{R}} K(x)dx = 1$ и $\int_{\mathbb{R}} K(x)x^k dx = 0$ для $k = 1, \ldots, n$.

Задача 4

T4 Оценка кросс-валидации параметра h определяется как точка минимума функции

$$\hat{J}(h) = \int \hat{p}_n^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{p}_{(-i)}(x_i),$$

где $\hat{p}_{(-i)}$ - оценка, построенная по всем значениям, кроме $i-$го (leave-one-out estimate). Для ядерной оценкой плотности $\hat{p}_n$, представьте функцию $\hat{J}(h)$ в виде функции, зависящей от выборки $x_1, .., x_n$ только через разности $x_i - x_j$, $i, j = 1, \dots, n$.

$$\int_{\mathbb{R}} (\hat{p}_n(x) - p(x))^2 dx = \int_{\mathbb{R}} (\hat{p}_n(x))^2 dx - 2\int_{\mathbb{R}} \hat{p}_n(x)\, p(x)\, dx + \int_{\mathbb{R}} p^2(x)\, dx \to \min$$

$= E[\hat{p}_n(X)]$, $X \sim p(x)$

$$E[\hat{p}_n(X)] = \frac{1}{n}\sum_{i=1}^{n} \hat{p}_{(-i)}(x_i)$$

$$\hat{J}(h) = \int_{\mathbb{R}} \frac{1}{n^2 h^2}\left(\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)\right)^2 dx - \frac{2}{n}\sum_{i=1}^{n}\sum_{j\neq i} \frac{1}{(n-1)h} K\left(\frac{x_i - x_j}{h}\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{(n-1)h} K\left(\frac{x_i - x_j}{h}\right) - \frac{2}{n}\cdot\frac{1}{(n-1)h} K(0)$$

$$\bullet\ \frac{1}{n^2h^2}\int_{\mathbb{R}} \sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{x - x_i}{h}\right) K\left(\frac{x - x_j}{h}\right) dx = \begin{bmatrix} u = \frac{x - x_i}{h} \\ x = uh + x_i \end{bmatrix} = \frac{1}{n^2h^2}\int_{\mathbb{R}} \sum_{i=1}^{n}\sum_{j=1}^{n} K(u)\, K\left(u + \frac{x_i - x_j}{h}\right) h\, du$$

$K(u) = K(-u)$

$$= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{\mathbb{R}} K(u)\, K\left(\frac{x_j - x_i}{h} - u\right) h\, du$$

$$f * g(x) = \int_{\mathbb{R}} f(x-u)\, g(u)\, du \Big\} = \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n} K * K\left(\frac{x_j - x_i}{h}\right)$$

$$\hat{J}(h) = \frac{1}{n^2 h}\sum_{i,j=1}^{n} K * K\left(\frac{x_j - x_i}{h}\right) - \frac{2}{n}\sum_{j=1}^{n} \frac{1}{(n-1)h} K\left(\frac{x_i - x_j}{h}\right) + \frac{2}{(n-1)h} K(0)$$

$h = \underset{h>0}{\operatorname{argmin}}\, \hat{J}(h)$ → "new"

"new": $\underset{h>0}{\operatorname{argmin}} \left\{ \frac{1}{n^2 h}\sum_{i,j=1}^{n} \left(K * K\left(\frac{x_j - x_i}{h}\right) - 2K\left(\frac{x_j - x_i}{h}\right)\right) + \frac{2}{(n-1)h} K(0) \right\}$